UPDATING OF VECTOR DATABASES AT THE INSTITUT CARTOGRAFIC DE CATALUNYA

M. Pla [a], A. Lleopart [a]

[a] ICC, Institut Cartogràfic de Catalunya, Parc de Montjuïc, E-08038 Barcelona, Spain
(maria.pla, anna.lleopart)@icc.cat

Commissions IV/VIII

KEY WORDS: Incremental updating, 2D and 2.5D databases, datamodel, generalization, web harvesting, GML, WFS.

ABSTRACT:

Since 1985 the ICC has been creating and updating topographic and thematic databases and deriving cartographic products using automatic symbolization and semi-automatic generalization. After describing the workflows used to update the ICC databases and the derived products, the paper describes the main aspects of ongoing projects for optimizing updating processes such as extending the data models for supporting incremental updating and versioning, the mechanisms for conveying updated data to the users, and the web based harvesting of information. In addition, the paper describes the prototype developed together with the Harbour Authority of Barcelona and the City of Barcelona for the collaborative updating of topographic data at large scale, and the project of updating the Street database that will be developed together with the municipalities.

1. INTRODUCTION

The Institut Cartogràfic de Catalunya (ICC) is a public law entity of the Generalitat de Catalunya, the Autonomous Government of Catalonia. From its creation in late 1982, the ICC has focused its efforts in the cartographic production in Catalonia either in promoting the studies to improve the production workflows.

Since its foundation the ICC has been collecting information for the generation and for the updating of topographic and thematic databases and deriving cartographic products using automatic symbolization and semi-automatic generalization. At scales up to 1:25.000, data is mainly collected using photogrammetric systems, and stored as 2.5D data models. Since the acquisition of the digital camera in 2004, these workflows are fully digital. Field survey is used to complete both the 1:1000 urban topographic and thematic databases, although thematic data is mainly provided by other departments of the government, municipalities or other public organizations.

2. CURRENT UPDATING WORKFLOWS

2.1 ICC databases

This section describes the characteristics and the current update workflows of the basic ICC topographic databases, the Topographic Database at 1:5000 scale and the Topographic Cartography at 1:1000 scale; and the Street database that is a thematic database.

2.1.1 Topographic Database at 1:5000 scale: The Topographic Database at 1:5000 scale (BT-5M) is the most detailed database that covers Catalonia. The data is compiled using photogrammetric systems according to a 2.5D data model and stored in DGN files from MicroStation. The accuracy is 1 meter RMS for X and Y and 1.5 meters RMS for Z. During the stereoplotting process all the features required to generate a digital terrain model (DTM) and a digital surface model (DSM) are compiled together with the topographic objects. The updating cycle is 5 years over all the country and more frequently over the most dynamic areas, located mainly in the coast.

The first compilation of the BT-5M started in 1985 and was completed in 1995. At that time, the underlying data model was based in 2.5D “spaghetti” vectors. It supported the generation of DTMs, but it was not designed for GIS. The next version of the model addressed this shortcoming. Moreover, it supported automatic generalization and 2.5D topographic objects. Updating started in 1996 using this new model and digital photogrammetric systems. Both changes were big enough to make the reuse of old data impossible. Therefore, the database was collected again from scratch with little or no use of existing data.

Since then, the data model has not changed and therefore the sole task is updating. All the objects of all layers (transportation, hydrography, settlement, terrain, etc) are updated from images of the most recent flight. In areas with large changes, all data is deleted and collected again (Figure 1). In areas with few changes (Figure 2) only the elements showing changes are collected again, and most of the information is maintained. New elements are added, non existing elements are removed and partial modifications are incorporated. Changes in the classification are also introduced, if needed.

Figure 1. At the left image the original data, at the centre the new image, at the right the updated information. The polygon indicates the area where original information has been completely replaced.

t

Figure 2. At the left image the original data, at the centre the new image, at the right the updated information. Circles indicate where original information has been partially modified.

Information must be changed in only two cases: when there is a real change on the object or when the object is erroneous. Other changes, for example based in aesthetical criteria or other, are not allowed.

2.1.2 Topographic cartography at 1:1 000 scale: The Topographic map series at 1:1000 scale (CT-1M) covers built-up areas and areas likely to become urban. The map is compiled using photogrammetric systems and completed by field survey. The file format is DGN. RMS is 20 centimeters in planimetry, and 25 centimeters in height. The data model is 2.5D, and it is designed to generate the DTM, the DSM, and to derive LOD 1 3D city models. The updating cycle is approximately 5 years.

All the objects of all layers are updated with images of the most recent flight or during the field survey. Usually most of the information is maintained and only the modified elements are collected again. It is also common that expanded areas, including completely new built-up areas or areas under planning, are added to the original area.

As in the BT-5M database, data is changed only when there is a real change on the object or when the object is erroneous. Other changes, for example based in aesthetical criteria, are not allowed.

2.1.3 Street database: The Street Database contains information of streets in all the built-up areas, housing developments and industrial areas of the country. The main feature type is the street centreline with attributes. Among others, the attributes are the feature identifier, the name of the street, the house numbers, at every side of the two ends of each street segment, and the ZIP code. Other feature types exist in the database, i.e. the street crossings and the quarters and municipal districts. They all are related together.

The main purpose of the database is standardizing and geocoding of addresses. It is used by the departments of the Catalan Government to standardise their address records (public equipments, emergency alerts, etc) and to locate them in the territory, and by the citizens through a web application

Most of the data is compiled on the field, but obtaining the information directly from the municipalities is also possible, especially in the updating process. The spatial reference for the geometry of the features is the Topographic Database of Catalonia 1:5 000.

The data is collected and maintained using ArcGIS. A software application handles add, modify, delete type of transactions. The process maintains the feature identifier, preserves the existing relations between features and keeps the history in the database. To fulfil these requirements, in addition to the feature identifier, every feature type in the database, whether it is spatial or not spatial, has a begin life span attribute and an end life span attribute. Whenever a feature is updated, the date of the life span end is set in its current version and a new version is created with the corresponding date of the life span beginning, so that the whole history of the feature is stored in the database.

New updating workflows are tested and will be implemented to facilitate the collaboration of the municipalities (see 3.1 and 3.2).

2.2 Derived products

In this section the updating workflows of two different products derived applying generalization from the BT-5M are described, the Topographic Map of Catalonia at 1:10 000 scale and the Topographic Database of Catalonia at 1:25 000 scale. Both are generated using DGN files.

2.2.1 Topographic Map of Catalonia at 1:10 000 scale (MT-10M): The Topographic Map of Catalonia at 1:10 000 scale (MT-10M) is just a map, not a database. It contains 2D data obtained by semiautomatic generalization from the BT-5M. Automatic processes include elimination of some objects, class aggregation, simplification of buildings, selection of altimetric points and selection and scaling of geographical names. Manual editing is applied to refine the automatic results, but because the small proportion between the original and the target scales, it is not required for most of the features and it can be performed in few hours. Updating cycle is the same that in the original database, 5 years over all the territory, and more frequent in the most dynamic areas.

In the design of the updating workflow of this product, the following options were considered: in the first one, based on incremental generalization, data is updated after generalization, modifying only the corresponding objects updated in the BT-5M; in the second one, all data is generalized again from the updated BT-5M without keeping any object of the old MT-10M. Although in the first option the number of changes is smaller, the amount of editing hours to adapt the new information to the unchanged information is higher. In the second option, because most of the generalization operations are automatic and the manual editing is quite fast, the number of hours is much smaller. Current workflow for MT-10M updating, as is showed in Figure 3, is based in the second option.

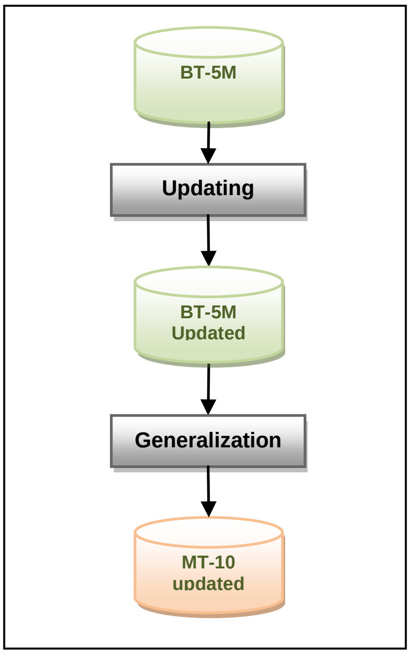

Figure 3. Updating workflow for the generalized MT-10M.

2.2.2 Topographic Database of Catalonia at 1:25 000 scale (MT-10M): The Topographic Database of Catalonia at 1:25 000 scale (BT-25M) is a 2.5D database produced by applying semiautomatic generalization to the BT-5M. The database is updated using stereoplotting of recent photogrammetric flights and it is completed with data extracted from thematic databases and, occasionally, with data collected on the field. Automatic generalization include the type of operations used for the MT-10M, plus the simplification of linear elements and the generalization of the elevation of the objects. Manual editing is used for generalization operations that can not be automatized such as typification, exaggeration, collapse, aggregation or refinement of the automatic results, and requires a quite high number of hours.

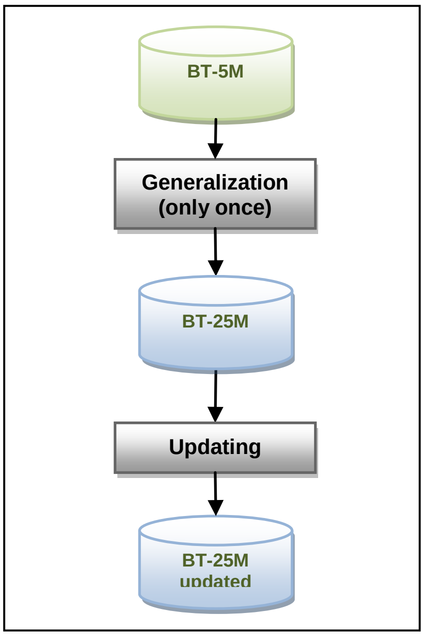

Figure 4. Updating workflow for the generalized BT-25M.

The BT-25M covers the 82% of the territory and will be completed during 2011. The database will be updated independently of the original database, because the updating cycle planned for the BT-25M will be approximately 2 years, much shorter than the updating cycle of the BT-5M. If the updating cycle of the original BT-5M could be reduced from 5 to 2 years, incremental generalization could be applied.

2.3 Problems and limitations

Main problems related with updating process during the production process of the ICC topographic databases are the cost of detecting changes and poor system's functionalities that makes it impossible to manage the requirements of the data model. Main problem related with distribution is the lack of information about the data that has been changed.

2.3.1 Detection of changes: In the process of updating the information using stereoplotting, part of the time is spent looking for the changes. In 2007, a project was developed at the ICC to systematically detect changes on the territory comparing SPOT images of different dates. The results were really useful for the purpose of the project, focussed on artificial coverages with surface larger than 1ha, but this resolution was too coarse for updating topographic data at larger scales. Therefore it would be interesting to test the results of the methodology with images of higher resolution. This should allow detecting changes in smaller areas or in linear elements. More difficulties can be found in the change detection for data collected on the field.

2.3.2 Limitations on the systems: The limitations on the systems used at the ICC to compile topographic databases cause a high cost of production and potential problems in the quality of the data. For example, because MicroStation cannot manage topological relationships, the topology of the topographic datasets has to be checked and rebuild again at every update.

Following table shows the productivity of the BT-5M, in hectares per hour, in the original data capture and in the data updating, distinguishing between stereoplotting process and editing process:

	Original capture (hes/h)	Updating (hes/h)
Steroplotting	20	29
Editing	44	48

Table 5. Productivity in original data capture and in updating process.

Increment of productivity in streoplotting process (45%) is not similar to the increment in editing process (18%). During stereoplotting only modified areas or elements must be manipulated, but during editing process, the complete topology has to be rebuilt again to warranty the requirements specified in the data model.

At this moment, new GIS tools with interactive topology based on ORACLE are under study, to substitute the data capture based on MicroStation DGN files. Main problem until now have been the lack of high productive photogrammetric systems based on GIS software.

2.3.3 Lack of information on modified data: Current data models for topographic databases used at ICC don't include information about versioning. In order to optimize the updating process and to facilitate the exploitation of the updated information to the users of the topographic databases, several attributes will be added to features, as the feature identifier, the begin life span attribute and the end life span attribute. The new tool based on GIS software will allow managing these new attributes with less effort than in the CAD tools used currently.

3. PROJECTS IN PROGRESS TO OPTIMIZE THE UPDATING WORKFLOWS

3.1 Prototype of collaborative updating

3.1.1 Background: This section describes the prototype developed by three organizations in Catalonia, the Institut Cartogràfic de Catalunya (ICC), the Port of Barcelona (Autoritat Portuària de Barcelona, APB), and the City of Barcelona (BCN) for building an automated infrastructure for sharing and harmonizing geospatial data, using Open Geospatial Consortium (OGC) standards, Geography Markup Language (GML) and Web Feature Specification (WFS).

3.1.2 Purpose and objectives: The common objective was the automation of the data transfer from the data producer to the data consumer using OGC standards to replace the old way of sending data changes via portable media or File Transfer Protocol (FTP), and the promotion of interoperability. As a result, access to the data of other organizations would be simpler and faster as the data changes can be harvested whenever the data consumer desires it.

In this project, APB and BCN play the role of data producers. They maintain their geographic data at 1:500 scale in SmallWorld and GeoMedia, respectively. APB provides data through a GeoServer WFS whereas BCN uses its GeoMedia WFS. APB is data consumer of BCN data to update its 1:500 database and ICC is data consumer of APB data, to update the Topographic Database of Catalonia at 1:5000 scale. It is worthwhile clarifying further that ICC is interested in harvesting the data changes incrementally rather than by taking complete snapshots of the APB database.

3.1.3 Process: To achieve this objective have been developed the following tasks:

- Mapping the data models.
- Creating GML schemas.
- Enhancing the data model to support the harvesting of data changes and to allow versioning of the data.
- Harvesting.

3.1.3.1 Mapping the data models: The creation of the mapping between the data models is one of the main challenges of inter-organizational geo-synchronization. Data experts of the respective organizations have to agree how to establish the relationships between features and attributes.
For the matching between APB and ICC models, the methodology has been to assign, to each feature and attribute of the APB data model, one feature and attribute of the ICC data model, whenever it exists. During matching no problems have been detected in geometry and topology. Semantic differences, due to different criteria in the classification of each model, have required some common sessions to understand both models, and to discuss and decide the matching.

The following table shows the number of APB and ICC features and the total features matched.

Layer	Features APB	Features APB matched	Features ICC	Features ICC matched
Orography	6	1	5	1
Buildings	28	8	19	5
Admin. Boundaries	8	-	-	-
Security	3	-	-	-
Equipment	28	15	7	7
Infraestructure	53	12	11	6
Names	4	1	1	1
Transport	28	3	5	3
Hydrography	2	2	5	2
External to the port	14	-	-	-
Maintenance	5	-	-	-
System	15	-	1	-
Sheet	3	-	-	-
Services	5	-	-	-
Total	202	42	54	25

Table 6. Number of features of each data model and number of matched features.

The encoding of the mapping is based in a XLST stylesheet, which is directly used by the harvesting process.

3.1.3.2 Creating the GML schemas: For the data to be retrieved and inserted through the WFS interface, a GML schema had to be defined for each data model.

APB used the software Feature Manipulation Engine (FME) to extract automatically their data from SmallWorld. The APB was able to configure FME to generate an additional version of the GML schema in which the features were disaggregated so that each feature had one geometric property, because they were better suited for mapping between APB and ICC features. ICC and BCN created automatically a new GML schema using GeoServer and GeoMedia, respectively.

3.1.3.3 Modelling the changes: In this project the data consumers are interested in harvesting incremental data updates that contain only data changes introduced by data producers since the previous harvest. It requires that the data producer make the changes (inserts, updates and deletes) available through the WFS interface, and that the data consumer have a way of determining what has changed. To allow this mechanism, the project participants had to enhance their data model expanding the GML schemas to include new feature properties that enabled the versioning of features. Three properties were added to enable capturing these feature events: operation, version start date, and version end date. The operation property indicates if the feature version was created as a result of insertion, update or deletion. The version start date denotes the date and time when the version was created. The version end date denotes the date and time when the version ceased to be current.

3.1.3.4 Harvesting the data changes through WFS: In this project the geo-synchronization uses the Pull model, i.e. the data consumer retrieves the data changes. In the Push model, the data producer sends the data to the data consumer as soon as the data changes. Both approaches have their pros and cons, but the Pull model was especially suitable in this project because the data consumer organizations didn't need a data update in real-time but only when they were ready to use it in their data maintenance processes. The figure 7 presents the workflow to retrieve data from the provider database and the subsequent transformation and insertion in the consumer database.

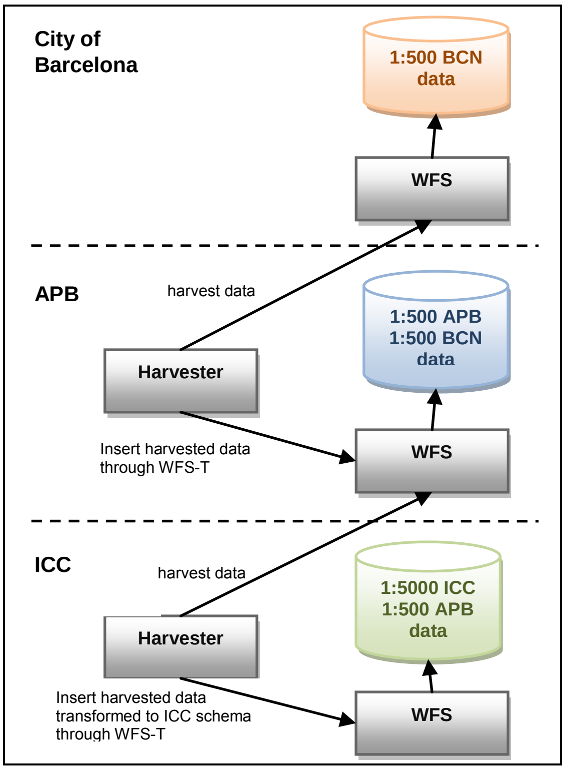

Figure 7. Overview of the mechanism to retrieve data between partners.

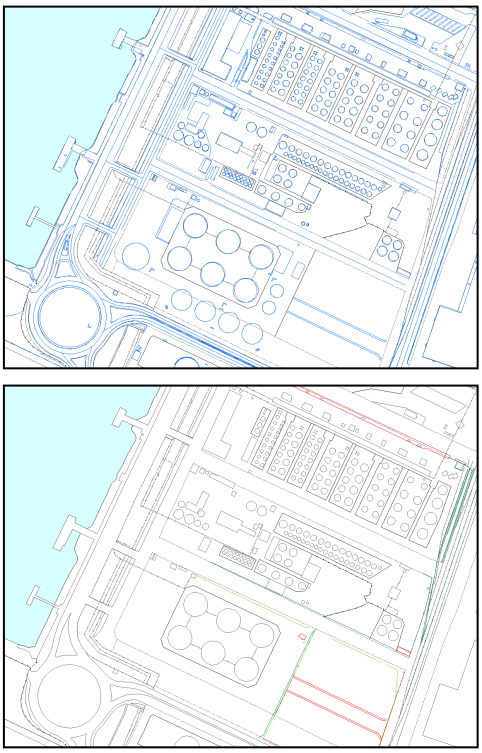

Figure 8. Examples of 1:500 APB harvested data on top of ICC BT-5M. At the top the result of the first harvesting (blue), at the bottom the result of the second one (red and green).

The initial APB harvest had fetched all the elements included in the matching, distributed as is showed in the following table. The 2nd harvest has retrieved only the elements modified since the 1st harvest.

Layers	1rst harvest	2rst harvest only changes
Equipments	3.956	269
Transport	494	1
Infraestructure	8.279	813
Buildings	3.245	117
Hydrography	8	7
Orography	40.613	5.107
Names	301	2
	56.896	6.316

Table 9. Number of elements harvested from APB database.

3.2 New updating workflows for the Street database

Municipalities have the most up-to-date information about the streets of their area. Therefore, they are the best source for the gathering of information to create and update the database. Unfortunately, in very few cases the data is available in such a way that it can be easily loaded in the database.

The analysis of the situation leads to establishing two possible scenarios, either in the creation or in the updating of the information: (1) the municipality has the information concerning the streets in a database; (2) the municipality has the information corresponding to streets in presentations other than a database (paper maps, CAD files, non geographical tables, other documents).

The current workflow for the first scenario is as follows:

- Mapping every municipality street data model (source) to the ICC data model (target).
- Development of a specific software application to transform from the source data model to the target data model, using the mapping table that has been previously drawn up.
- Reception of the updates from the municipality.
- Automatic transformation of the mapped data.
- Analysis of the data that has not been possible to extract automatically and its manual editing.
- Insertion of the updated data in the database.

In the second scenario, the workflow comes down to the interpretation of the available information, manual editing and insertion in the database.

In order to improve these workflows for the updating process and increase the collaboration between the municipalities and ICC, new workflows are being designed for the scenarios described previously. They are briefly described below.

3.2.1 Web based incremental harvesting of the data changes: The new workflow for the first scenario substitutes the transformation process of the current workflow by a web based new one which is an actual application of the prototype of collaborative updating exposed in the previous chapter. It is important to note that the improvement lies in the substitution of the transfer of the updated data by ftp or DVD, and the subsequent transformation, by the request of automatic harvesting of the changes from the source database. The defined data mapping is encoded in a XSLT stylesheet so that the data is transformed from the source to the target data model, by the harvester, applying the stylesheet on the source GML data retrieved from the WFS.

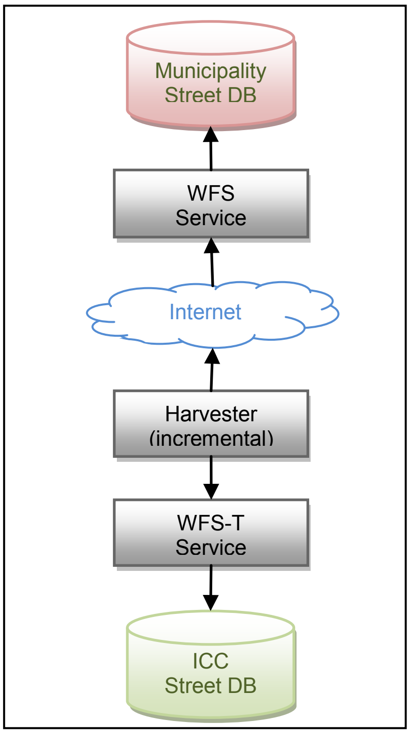

Figure 10. ICC Street database updating through incremental harvesting of the data changes of the Municipality Street database.

3.2.2 Web based updating of the database: In the second scenario, a completely new workflow is devised. Provided that the municipality does not have the data in a database, a web application will be put in place so that the municipality could introduce the proposed changes directly in the target database through WFS-T. Afterwards, the changes will be checked and consolidated in the database.

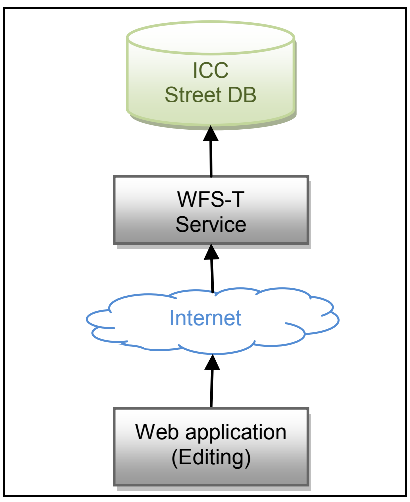

Figure 11. Update of the ICC Street database by the municipality using a web application.

4. CONCLUSIONS

The results of the prototype of collaborative updating demonstrates that the implementation of infrastructures based on OGC, GML and WFS standards to share geographic data can facilitate the distributed collaborative update between ICC and other organizations. Its application on the updating of the Street database will optimize the complete updating process and, moreover, will improve the quality of the information. Furthermore, it can be applied in the updating of some thematic information of the topographic databases.

Another aspect to consider in the case of the ICC topographic databases is the improvement of the data model, with the inclusion of unique identifiers and attributes to keep the feature versions in order to have historic information and to allow incremental updating. The migration to GIS tools will also improve the database management, will enhance the data quality during data capture, and will spread the exploitations done by producers and users.

5. REFERENCES

Milanović, A., Guimet, J., Rodellas, E., Bolívar, M.A., 2010 *Interorganizational Geo-Synchronization Using OGC Technologies to Share and Harmonize Data in Catalonia*. To be presented in INSPIRE Conference June 2010.